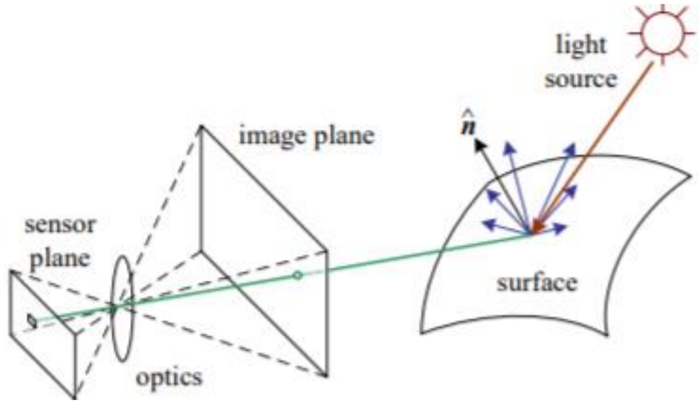
Image Formation

- The study of image formation encompasses the radiometric and geometric processes by which 2D images of 3D objects are formed.
- In the case of digital images, the image formation process also includes analog to digital conversion and sampling.
- Apart from geometric features, image formation also depends on discrete color and intensity values.
- It needs to know the lighting of the environment, camera optics, sensor properties, etc.

Photometric Image Formation

Figure below gives a simple explanation of image formation.



- The light from a source is reflected on a particular surface.
- A part of that reflected light goes through an image plane that reaches a sensor plane via optics.

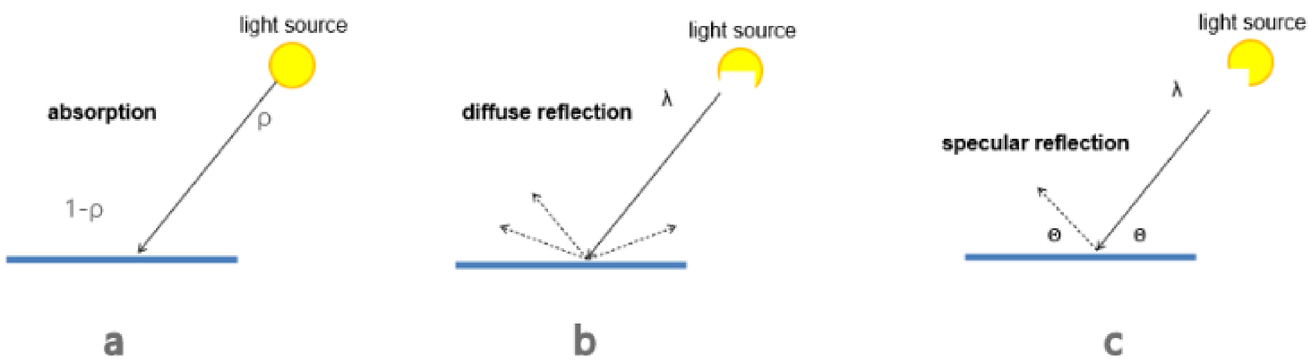
Some factors that affect image formation are:

- The strength and direction of the light emitted from the source.
- The material and surface geometry along with other nearby surfaces.
- Sensor Capture properties

Reflection and Scattering

Images cannot exist without light. Light sources can be a point or an area light source. When the light hits a surface, three major reactions might occur-

1. Some light is absorbed. That depends on the factor called ρ (albedo). Low ρ of the surface means more light will get absorbed.
2. Some light gets reflected diffusively, which is independent of viewing direction. It follows **Lambert's cosine law** that the amount of reflected light is proportional to cos(θ). E.g., cloth, brick.
3. Some light is reflected specularly, which depends on the viewing direction. E.g., mirror.



- Apart from the above models of reflection, the most common model of light scattering is the **Bidirectional Reflectance Distribution Function (BRDF).**
- It gives the measure of light scattered by a medium from one direction into another.
- The scattering of the light can determine the topography of the surface — smooth surfaces reflect almost entirely in the specular direction, while with increasing roughness the light tends to diffract into all possible directions.
- Eventually, an object will appear equally bright throughout the outgoing hemisphere if its surface is perfectly diffuse (i.e., Lambertian). Owing to this, BRDF can give valuable information about the nature of the target sample.

Color

From a viewpoint of color, we know visible light is only a small portion of a large electromagnetic spectrum.

Two factors are noticed when a colored light arrives at a sensor:

- Colour of the light
- Colour of the surface

Bayer Grid/Filter is an important development to capture the color of the light. In a camera, not every sensor captures all the three components (RGB) of light. Inspired by human visual preceptors, Bayers proposed a grid in which there are 50% green, 25 % red, and 25% blue sensors.

Demosaicing algorithm is then used to obtain a full-color image where the surrounding pixels are used to estimate the values for a particular pixel.

There are many such color filters that have been developed to sense colors apart from Bayer Filter.

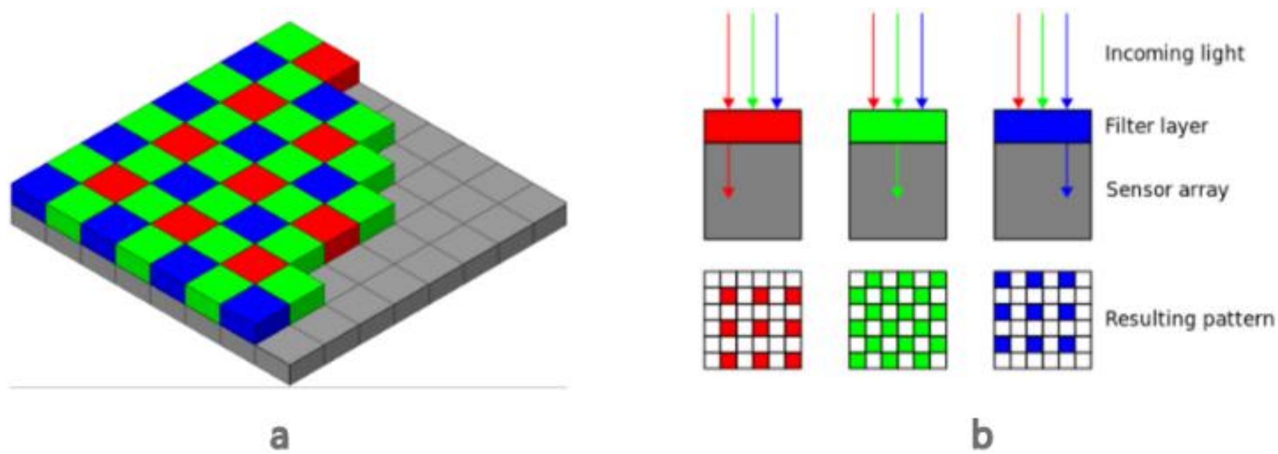
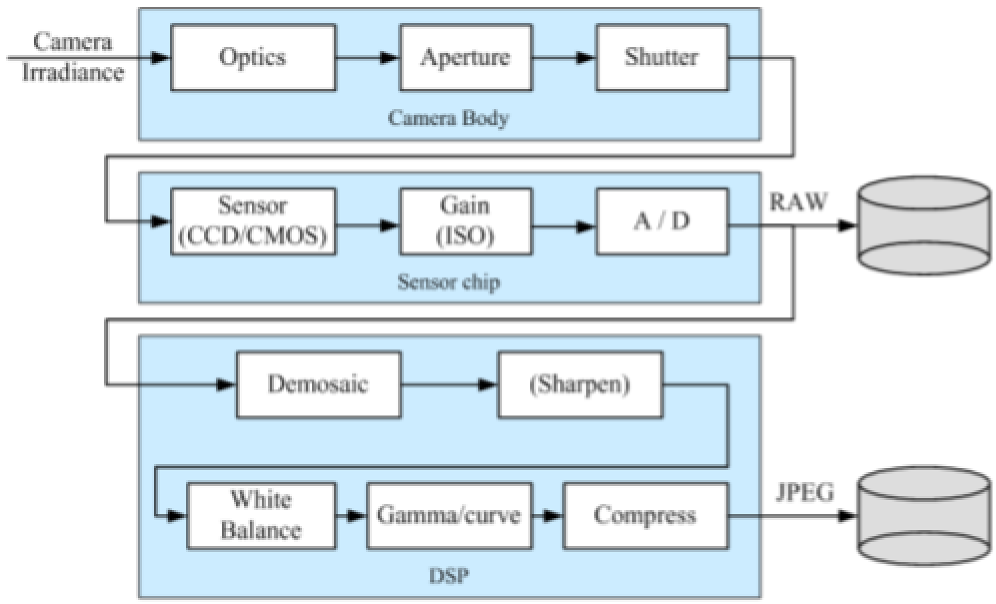


Image sensing Pipeline (The digital camera)

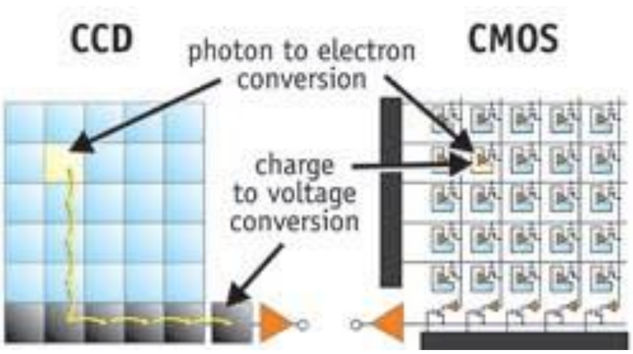
- The light originates from multiple light sources, gets reflected on multiple surfaces, and finally enters the camera where the photons are converted into the (R, G, B) values that we see while looking at a digital image.



- In a camera, the light first falls on the lens (optics).
- Following that is the aperture and shutter which can be specified or adjusted.
- Then the light falls on sensors which can be charged coupled device (CCD) or complementary metal-oxide-semiconductor (CMOS), then the image is obtained in an analog or digital form and we get the raw image.

CCD vs CMOS

- The camera sensor can be CCD or CMOS.
- In charged coupled device (CCD). A charge is generated at each sensing element and this photogenerated charge is moved from pixel to pixel and is converted into a voltage at the output node. Then an analog to digital converter (ADC) converts the value of each pixel to a digital value.
- The complementary metal-oxide-semiconductor (CMOS) sensors work by converting charge to voltage inside **each** element as opposed to CCD which accumulates the charge. CMOS signal is digital and therefore does not need ADC. CMOS is widely used in cameras in the current times.



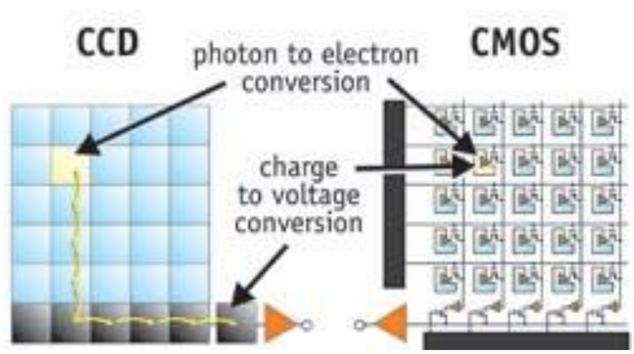
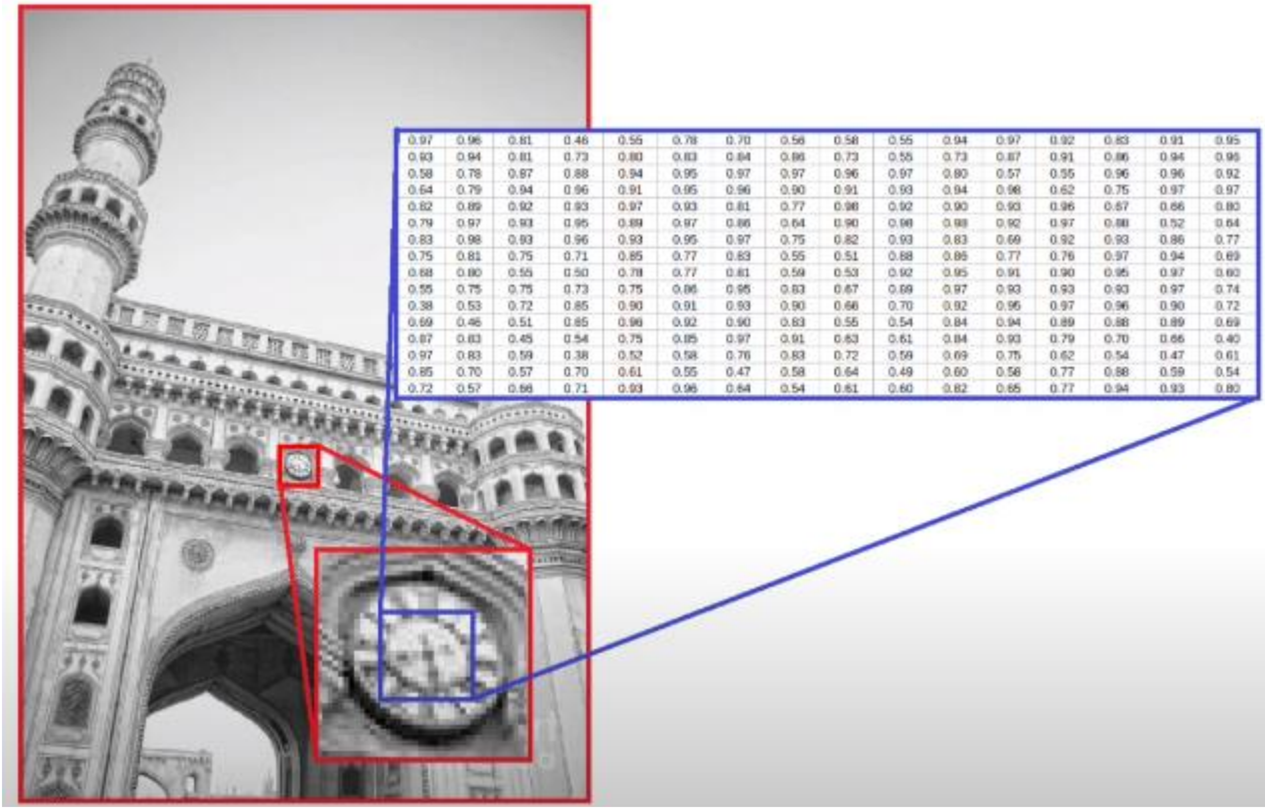


Image as a matrix

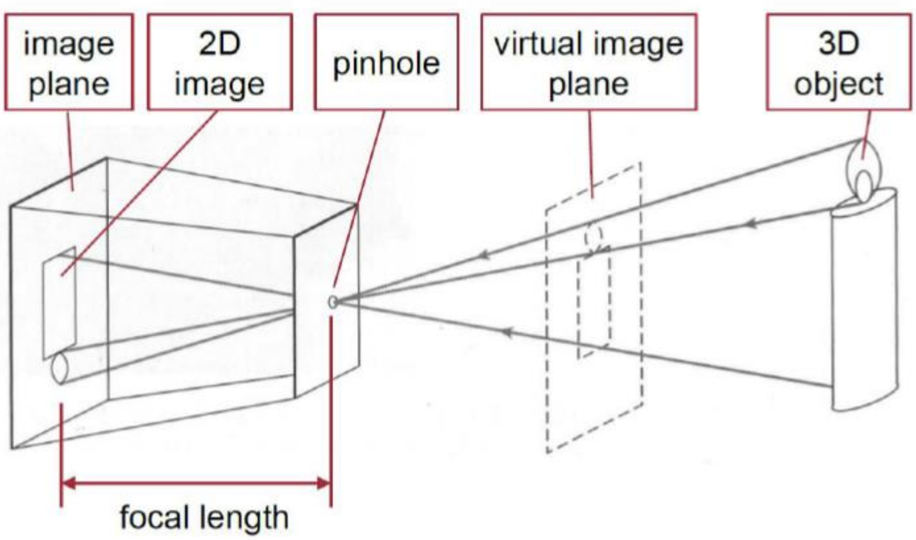
The simplest way to represent the image is in the form of a matrix.



- It is commonly seen that people use up to a byte to represent every pixel of the image.
- This means that values between 0 to 255 represent the intensity for each pixel in the image where 0 is black and 255 is white.
- For every color channel in the image, one such matrix is generated. In practice, it is also common to normalize the values between 0 and 1 (as done in the example in the figure above).

Camera models

- To understand how vision might be modeled computationally and replicated on a computer, we need to understand the image acquisition process.



Pinhole Camera

Pinhole camera model

- The pinhole camera is the simplest, and the ideal, model of camera function.
- It has an infinitesimally small hole through which light enters before forming an inverted image on the camera surface facing the hole.
- It this we place image plane *between* the focal point of the camera and the object, so that the image is not inverted.
- This mapping of three dimensions onto two, is called a perspective projection, and perspective geometry.

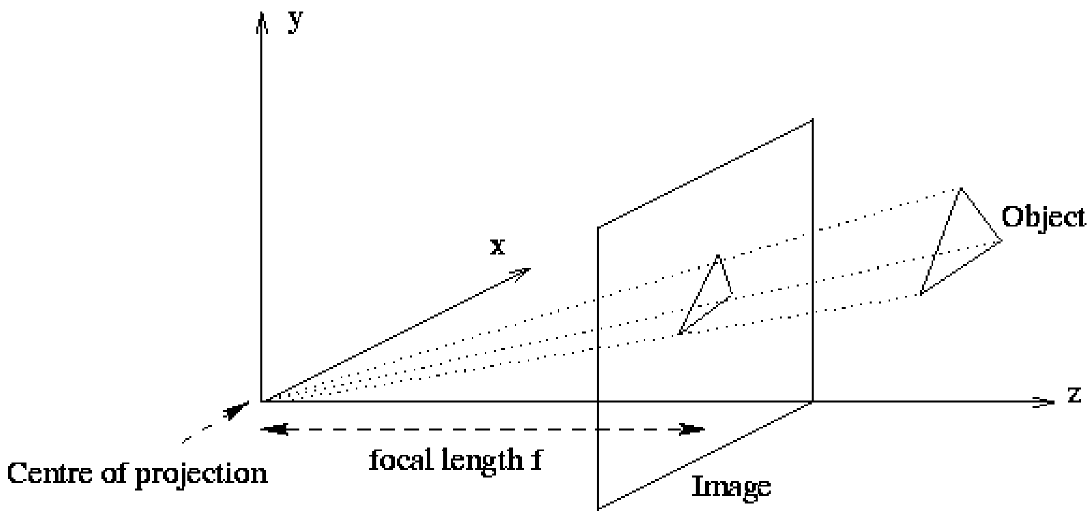


Figure: Perspective projection in the pinhole camera model.

Perspective geometry

- Euclidean geometry is a special case of perspective geometry, and the use of perspective geometry in computer vision makes for a simpler and more elegant expression of the computational processes that render vision possible.

- A perspective projection is the projection of a three-dimensional object onto a two-dimensional surface by straight lines that pass through a single point.

- Simple geometry shows that if we denote the distance of the image plane to the **centre of projection by *f*,** then the image coordinates (x_i, y_i) are related to the object coordinates (x_o, y_o, z_o) by

$$x_i = \frac{f}{z_o} x_o$$

$$y_i = \frac{f}{z_o} y_o$$

  and

- These equations are non-linear. They can be made linear by introducing *homogeneous transformations*, which is effectively just a matter of placing the Euclidean geometry into the perspective framework.

- Each point (x, y, z) in three-space is mapped onto a line in four-space given by (wx, wy, wz, w), **where *w* is a dummy variable that sweeps out the line ($w \neq 0$)**.

- In homogeneous coordinates, the perspective projection onto the plane is given by

$$\begin{bmatrix} x_i \\ y_i \\ z_i \\ w_i \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & f & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_o \\ y_o \\ z_o \\ w_o \end{bmatrix} .$$

- The *projective plane* is used to model the image plane.
- A point in the plane is represented by a 3-vector (x_1, x_2, x_3) of numbers not all zero.
- They define a vector **x** up to a scale factor.
- A line **l** is also defined by a triplet of numbers (u_1, u_2, u_3), not all zero, and satisfies the equation

  $u_1x + u_2y + u_3 = 0$.

A point on a line is given by the relations

$$\mathbf{l}.\mathbf{x} = 0 \quad \text{or} \quad \mathbf{l}^T\mathbf{x} = 0 \quad \text{or} \quad \mathbf{x}^T\mathbf{l} = 0.$$

Two points define a line by the equation

$$\mathbf{l} = \mathbf{p} \wedge \mathbf{q}$$

where $\wedge$ denotes the vector product.

Likewise, two lines define a point by the equation

$$\mathbf{x} = \mathbf{l} \wedge \mathbf{m}$$

- This duality between lines and points in the image plane is often exploited in homogeneous notation.

As well, a vector product can also be written in matrix notation, by writing the vector as a skew-symmetric matrix. Thus,

$$\mathbf{v} \wedge \mathbf{x} = \begin{bmatrix} 0 & -v_z & v_y \\ v_z & 0 & -v_x \\ -v_y & v_x & 0 \end{bmatrix} \mathbf{x}.$$

projective space is used as a model for Euclidean 3-space. Here, points and planes are represented by quadruplets of numbers not all zero, and are duals of each other.

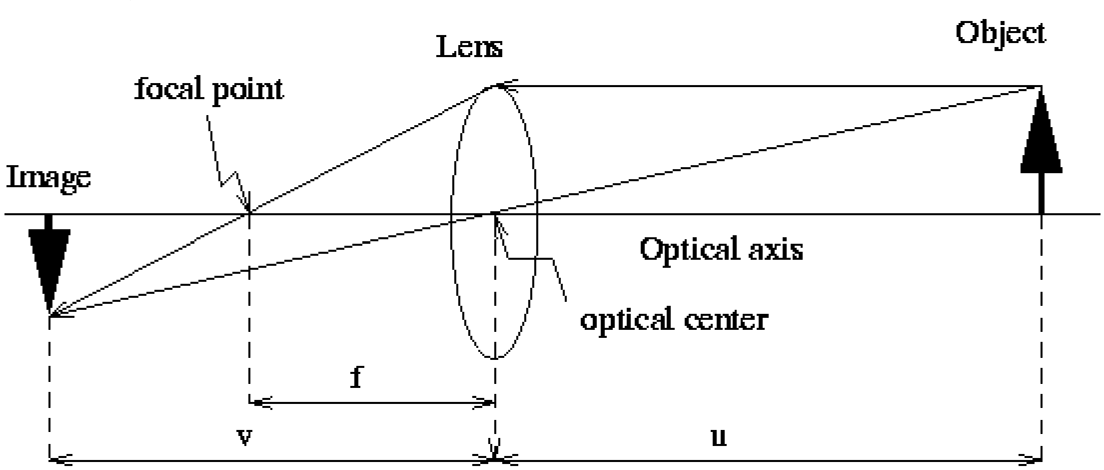


Figure: Simple lens model.

Pinhole Camera Coordinate System

The simplest camera model is pinhole model which decribes the mathematical relationship of the projection of points in 3d-space onto a image plane. Let the centre of projection be the origin of a Euclidean coordinate system, and the plane $Z = f$, which is called the *image plane* or *focal plane*. Under pinhole camera model, a point in space with coordinates $(X, Y, Z)^T$ is mapped to the point on the image plane $(\frac{fX}{Z}, \frac{fY}{Z}, f)^T$ using triangles as shown in Figure 1. Ignoring the final image coordinate, the central projection mapping from 3d world space to 2d image coordinates is,

$$(X, Y, Z)^T \longrightarrow (\frac{fX}{Z}, \frac{fY}{Z})^T \tag{1}$$

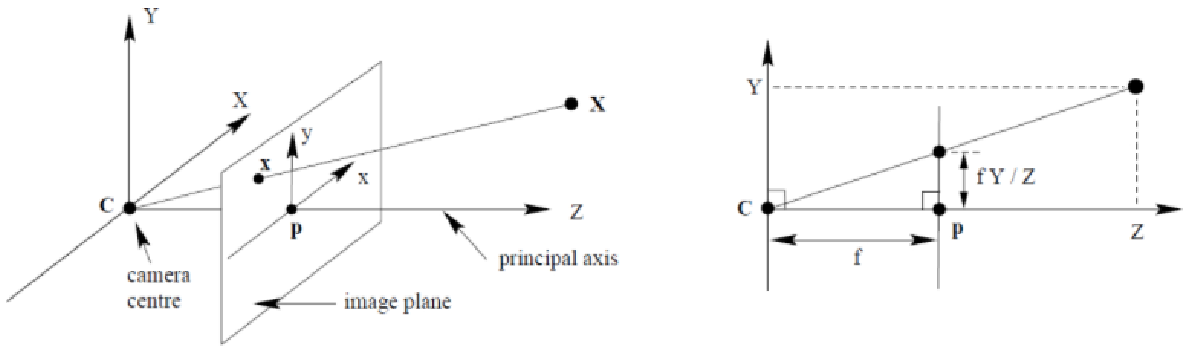


Figure 1. Pinhole camera geometry. The centre of projection is called the *camera centre* or the *optical centre*. The line from the camera centre perpendicular to the image plane is called the *principal axis* or *principal ray*. The point where the principal axis meets the image plane is called the *principal point*. The plane through the camera centre parallel to the image plane is called the *principal plane of the camera*. **C** is the camera centre and **p** the principal point. The camera centre is here placed at the coordinate origin [Hartley and Zisserman, 2003].

Assuming the world and image points are represented in homogeneous coordinates, then central projection can simply expressed as a linear mapping between their homogeneous coordinates in terms of matrix multiplication by,

$$\begin{bmatrix} fX \\ fY \\ Z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_{cam} \\ Y_{cam} \\ Z_{cam} \\ 1 \end{bmatrix} \tag{2}$$

Principal point offset: In theory the origin of coordinates in the image plane assumed to be at the principal point. This may not be true in practice, hence, the Eq. (2) is express as,

$$\begin{bmatrix} fX + Zp_x \\ fY + Zp_y \\ Z \end{bmatrix} = \begin{bmatrix} f_x & 0 & p_x & 0 \\ 0 & f_y & p_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_{cam} \\ Y_{cam} \\ Z_{cam} \\ 1 \end{bmatrix} \tag{3}$$

First matrix in the right side of Eq. (3) called camera *calibration matrix*, usually expressed by K. For added generality, the calibration matrix can be express as,

$$K = \begin{bmatrix} \alpha_x & s & p_x \\ 0 & \alpha_y & p_y \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where s is referred to as the skew parameter which is zero for most of the cameras. f_x and f_y where $alpha_x = fm_x$ and $\alpha_y = fm_y$ represent the focal length of the camera in terms of pixel dimensions in the *x-axis* and the *y-axis* respectively, and (p_x, p_y) is coordinate of the principal point.

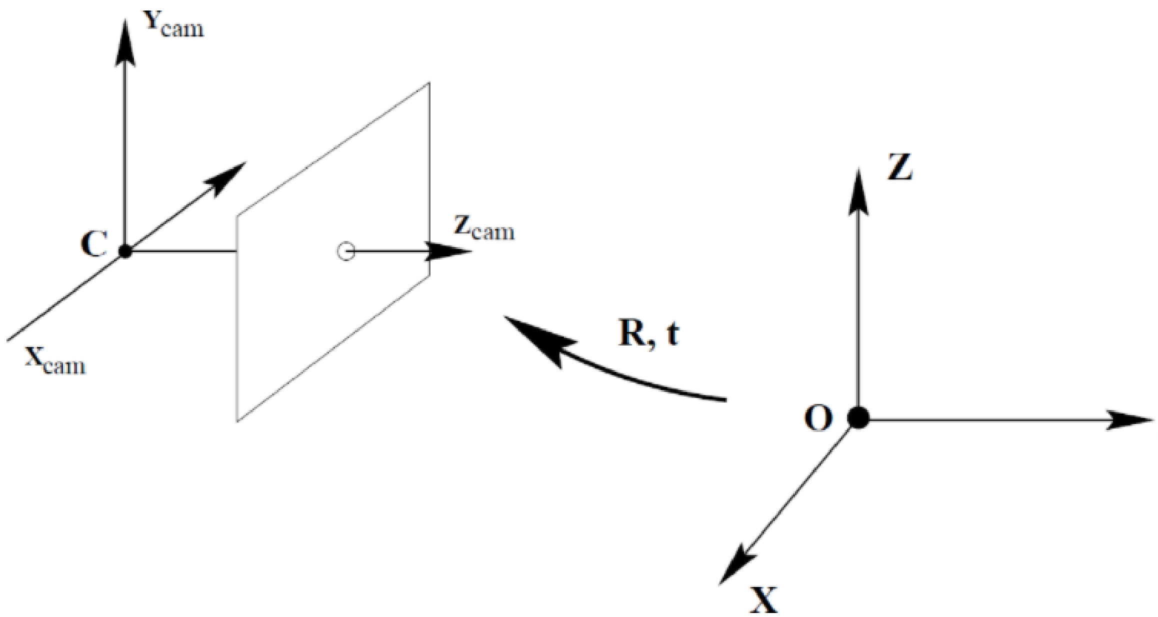


Figure 2: The Euclidean transformation between the world and camera coordinate frames [Hartley and Zisserman, 2003].

Camera rotation and translation: The Subscripts *cam* in above equations is to emphasize that the points are represented as *camera coordinate frame*. In general, points in space is determined by the *world coordinate frame*. The camera coordinate and world coordinate frames are related by *rotation* and *translation*. As it is shown in Figure 2, if $\mathbf{X} = (X, Y, Z, 1)^T$ is the coordinate of the point in the world coordinates, then $\mathbf{X}_{cam}$ is transformed by,

$$\mathbf{X}_{cam} = [R \quad t]\mathbf{X} \tag{5}$$

where **R** is 3×3 rotation matrix and **t** is 3×1 translation matrix.

Putting eveything together the formula for general mapping of pinhole camera in world coordinate frame **x** is defined by,

$$\mathbf{x} = \mathbf{K}[\mathbf{R} \quad \mathbf{t}]\mathbf{X} \tag{6}$$

So that, the general pinhole camera matrix, **P**, can be represented by,

$$\mathbf{P} = \mathbf{K}[\mathbf{R} \quad \mathbf{t}] = \begin{bmatrix} f_x & 0 & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_1 & r_2 & r_3 & t_1 \\ r_4 & r_5 & r_6 & t_2 \\ r_7 & r_8 & r_8 & t_3 \end{bmatrix} \tag{7}$$

and it has 9 degrees of freedom,

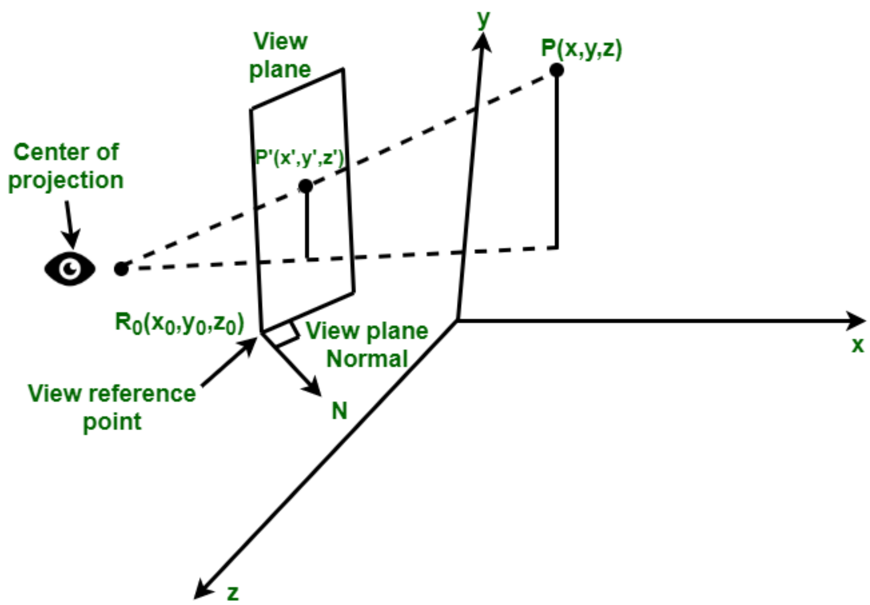
1. Three for **K**, namely, f, p_x, and p_y
2. Three for rotation matrix **R**
3. Three for translation **t**

Internal camera parameters, K , show the internal orientation of the camera and it is fixed.

External parameters, R and t show camera orientation and position to a world coordinate system.

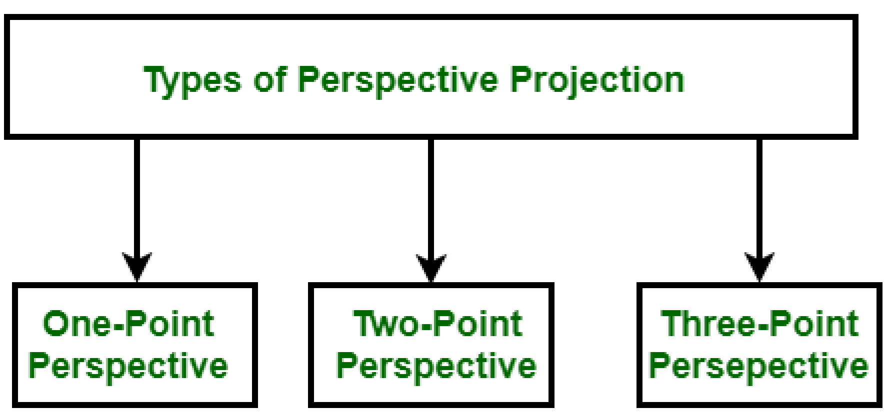
Perspective Projection and its Types

- In Perspective Projection the **center of projection** is at finite distance from **projection plane**.
- This projection produces realistic views but does not preserve relative proportions of an object dimensions.
- Projections of distant object are smaller than projections of objects of same size that are closer to projection plane. The perspective projection can be easily described by following figure:

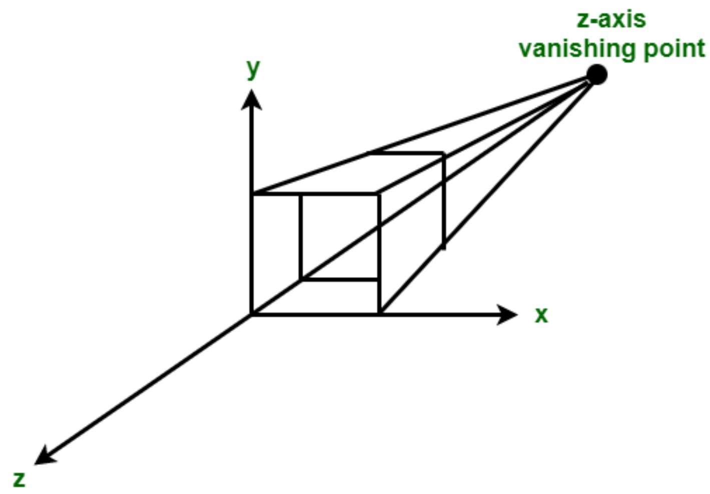


1. **Center of Projection –** It is a point where lines or projection that are not parallel to projection plane appear to meet.
2. **View Plane or Projection Plane –** The view plane is determined by :
   - View reference point $R_0(x_0, y_0, z_0)$
   - View plane normal.
3. **Location of an Object –** It is specified by a point P that is located in world coordinates at (x, y, z) location. The objective of perspective projection is to determine the image point P' whose coordinates are (x', y', z')

Types of Perspective Projection : Classification of perspective projection is on basis of vanishing points (It is a point in image where a parallel line through center of projection intersects view plane.). We can say that a vanishing point is a point where projection line intersects view plane.

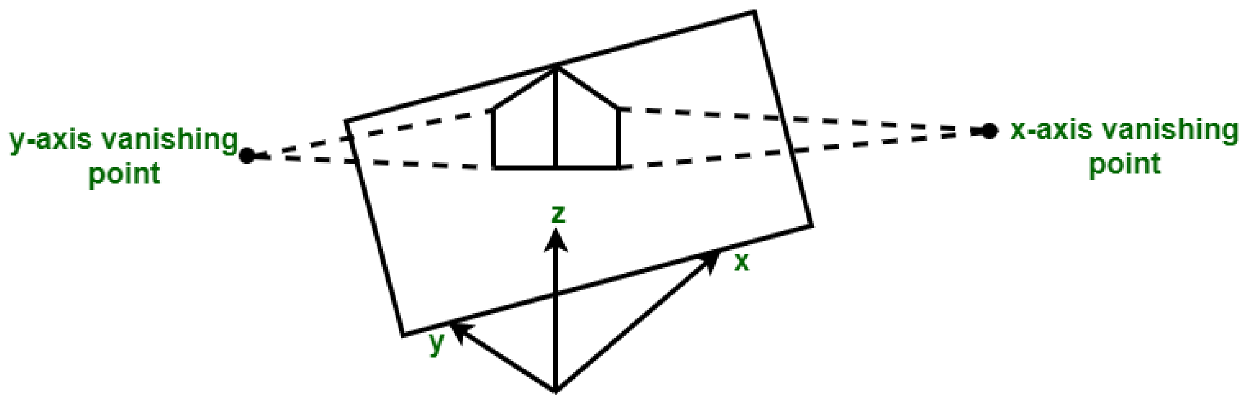


- **One Point Perspective Projection** – One point perspective projection occurs when any of principal axes intersects with projection plane or we can say when projection plane is perpendicular to principal axis.



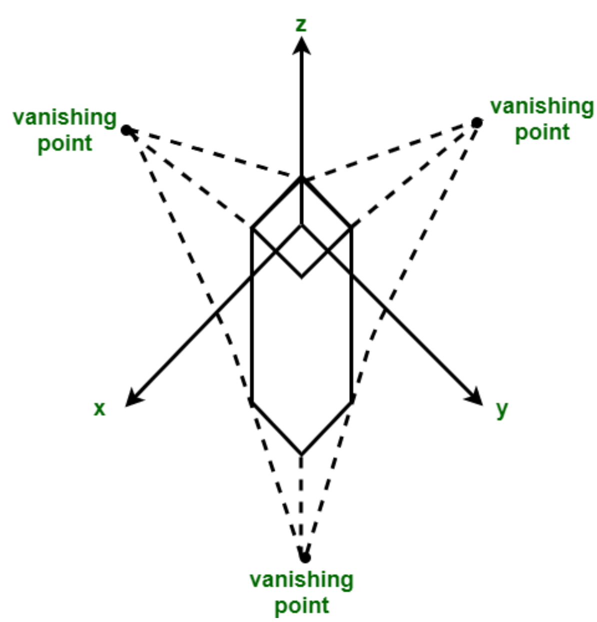
In the above figure, z axis intersects projection plane whereas x and y axis remain parallel to projection plane.

- **Two Point Perspective Projection** –Two point perspective projection occurs when projection plane intersects two of principal axis.



In the above figure, projection plane intersects x and y axis wheres z axis remains parallel to projection plane.

- **Three Point Perspective Projection** –Three point perspective projection occurs when all three axis intersects with projection plane. There is no any principle axis which is parallel to projection plane.

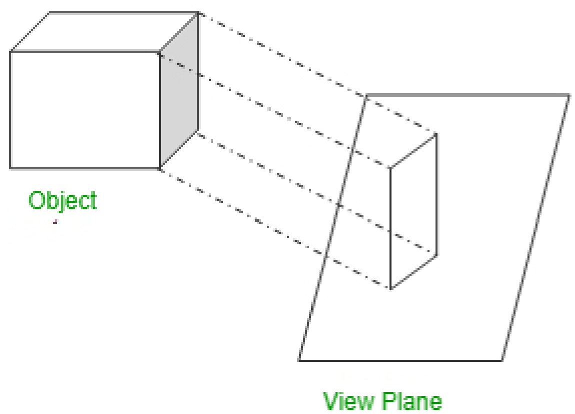


Application of Perspective Projection : The perspective projection technique is used by artists in preparing drawings of three-dimensional objects and scenes.

Difference between Parallel and Perspective Projection in Computer Graphics

1. Parallel Projection : Parallel projections are used by architects and engineers for creating working drawing of the object, for complete representations require two or more views of an object using different planes.

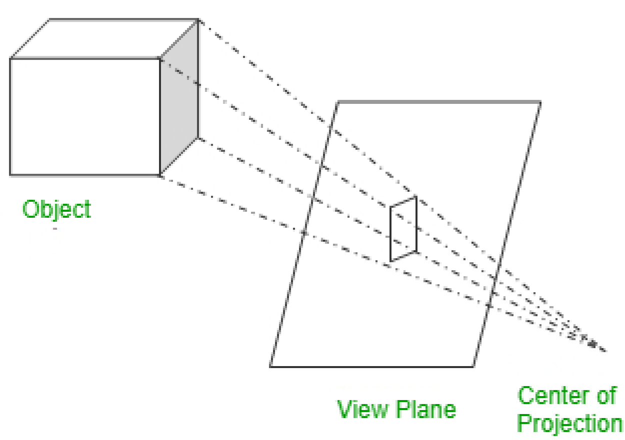
Parallel Projection use to display picture in its true shape and size. When projectors are perpendicular to view plane then is called orthographic projection. The parallel projection is formed by extending parallel lines from each vertex on the object until they intersect the plane of the screen. The point of intersection is the projection of vertex.



2. Perspective Projection : Perspective projections are used by artist for drawing three-dimensional scenes.

In Perspective projection lines of projection do not remain parallel. The lines converge at a single point called a center of projection. The projected image on the screen is obtained by points of intersection of converging lines with the plane of the screen. The image on the screen is seen as of viewer's eye were located at the centre of projection, lines of projection would correspond to path travel by light beam originating from object.

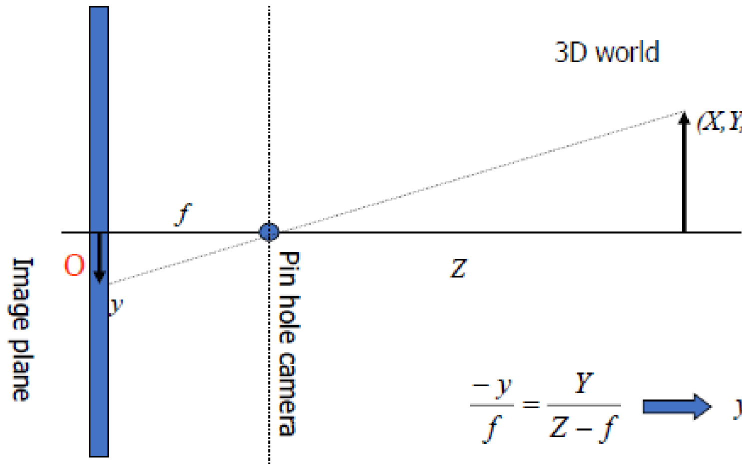
Two main characteristics of perspective are vanishing points and perspective foreshortening. Due to foreshortening object and lengths appear smaller from the center of projection. More we increase the distance from the center of projection, smaller will be the object appear.



Difference Between Parallel Projection And Perspective Projection :

SR.NO	Parallel Projection	Perspective Projection
1	Parallel projection represents the object in a different way like telescope.	Perspective projection represents the object in three dimensional way.
2	In parallel projection, these effects are not created.	In perspective projection, objects that are far away appear smaller, and objects that are near appear bigger.
3	The distance of the object from the center of projection is infinite.	The distance of the object from the center of projection is finite.
4	Parallel projection can give the accurate view of object.	Perspective projection cannot give the accurate view of object.
5	The lines of parallel projection are parallel.	The lines of perspective projection are not parallel.
6	Projector in parallel projection is parallel.	Projector in perspective projection is not parallel.
7	Two types of parallel projection : 1.Orthographic, 2.Oblique	Three types of perspective projection: 1.one point perspective, 2.Two point perspective, 3. Three point perspective,
8	It does not form realistic view of object.	It forms a realistic view of object.

Perspective Projection(Origin at image center)



$$\frac{-y}{f} = \frac{Y}{Z-f} \Rightarrow y = -f\frac{Y}{Z-f} \Rightarrow y = \frac{fY}{f-Z}$$

$$x = \frac{fX}{f-Z}$$

Homogeneous coordinates

$$(x, y) \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

homogeneous image coordinates

$$(x, y, z) \Rightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

homogeneous scene coordinates

Converting *from* homogeneous coordinates

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow (x/w, y/w)$$

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \Rightarrow (x/w, y/w, z/w)$$

Perspective Projection

Projection is a matrix multiply using homogeneous coordinates:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ -z/d \end{bmatrix} \Rightarrow (-d\frac{x}{z}, \quad -d\frac{y}{z})$$

divide by third coordinate

This is known as **perspective projection**

- The matrix is the **projection matrix**
- Can also formulate as a 4x4 (today's reading does this)

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1/d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \\ -z/d \end{bmatrix} \Rightarrow (-d\frac{x}{z}, \quad -d\frac{y}{z})$$

Perspective Projection

How does scaling the projection matrix change the transformation?

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ -z/d \end{bmatrix} \Rightarrow (-d\frac{x}{z}, \quad -d\frac{y}{z})$$

$$\begin{bmatrix} -d & 0 & 0 & 0 \\ 0 & -d & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} -dx \\ -dy \\ z \end{bmatrix} \Rightarrow (-d\frac{x}{z}, \quad -d\frac{y}{z})$$

Perspective Projection Example

1. Object point at (10, 6, 4), d=2

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/2 & 0 \end{bmatrix} \begin{bmatrix} 10 \\ 6 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 10 & 6 & -2 \end{bmatrix}$$

$$\Rightarrow \quad x' = -5, \; y' = -3$$

2. Object point at (25, 15, 10)

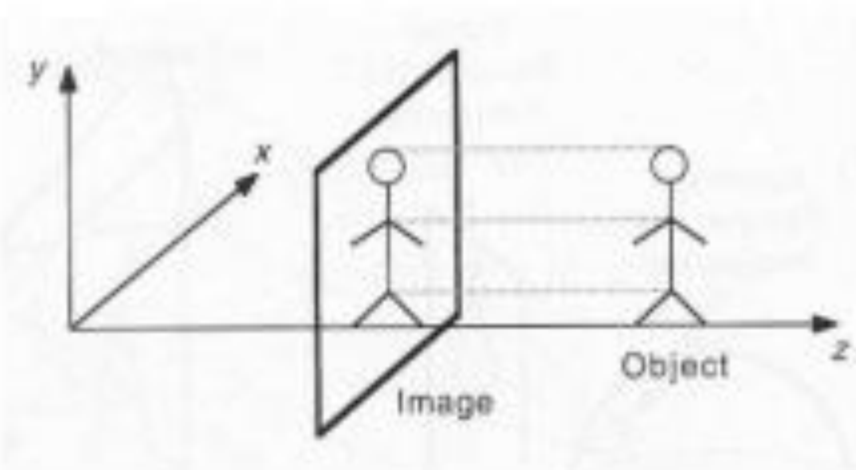
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/d & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/2 & 0 \end{bmatrix} \begin{bmatrix} 25 \\ 15 \\ 10 \\ 1 \end{bmatrix} = \begin{bmatrix} 25 & 15 & -5 \end{bmatrix}$$

$$\Rightarrow \quad x' = -5, \; y' = -3$$

Orthographic Projection

- It is the projection of a 3D object onto a plane by a set of parallel rays orthogonal to the image plane.

- It is the limit of perspective projection as $f -> \infty$ (i.e., $f/Z -> 1$)



orthographic proj. eqs: $x = X$, $y = Y$ (drop Z)

- Using matrix notation:

$$\begin{bmatrix} x_h \\ y_h \\ z_h \\ w \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

- Verify the correctness of the above matrix (homogenize using $w=1$):

$$x = \frac{x_h}{w} = X \quad y = \frac{y_h}{w} = Y$$

• Properties of orthographic projection

- Parallel lines project to parallel lines.

- Size does not change with distance from the camera.

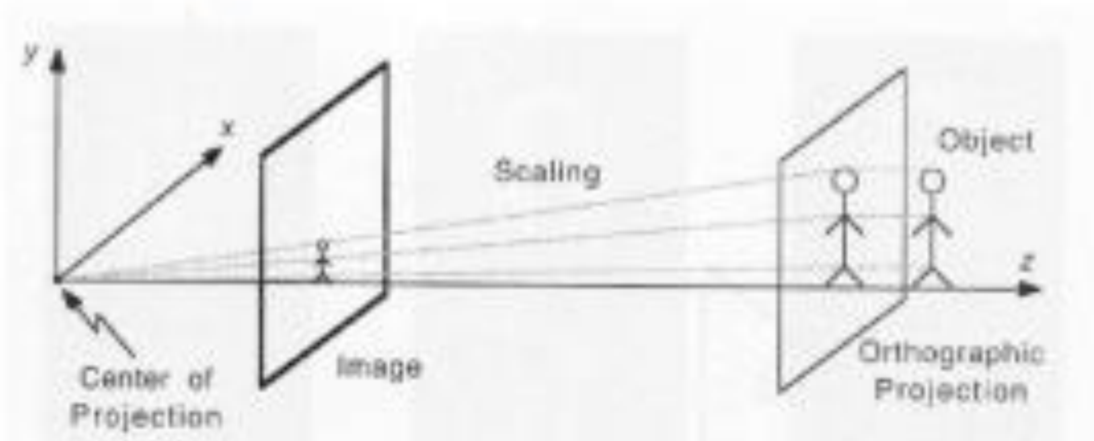
Weak Perspective Projection

- Perspective projection is a non-linear transformation.

- We can approximate perspective by scaled orthographic projection (i.e., linear transformation) if:

(1) the object lies <u>close</u> to the optical axis.

(2) the object's dimensions are <u>small</u> compared to its average distance $\bar{Z}$ from the camera (i.e., $\delta z < \bar{Z}/20$)



weak perspective proj. eqs: $x = \frac{Xf}{Z} = \frac{Xf}{\bar{Z}}$ $\quad y = \frac{Yf}{Z} = \frac{Yf}{\bar{Z}}$ (drop Z)

- The term $\frac{f}{\bar{Z}}$ is a scale factor now (e.g., every point is scaled by the same factor).

- Using matrix notation:

$$\begin{bmatrix} x_h \\ y_h \\ z_h \\ w \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \bar{Z} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

- Verify the correctness of the above matrix (homogenize using $w = \bar{Z}$):

$$x = \frac{x_h}{w} = \frac{fX}{\bar{Z}} \quad y = \frac{y_h}{w} = \frac{fY}{\bar{Z}}$$